NONCONVENTIONAL POISSON LIMIT THEOREMS

YURI KIFER

INSTITUTE OF MATHEMATICS
HEBREW UNIVERSITY
JERUSALEM, ISRAEL

Abstract. The classical Poisson theorem says that if $\xi_1, \xi_2, ...$ are i.i.d. 0–1 Bernoulli random variables taking on 1 with probability $p_n \equiv \lambda/n$ then the sum $S_n = \sum_{i=1}^n \xi_i$ is asymptotically in n Poisson distributed with the parameter λ. It turns out that this result can be extended to sums of the form $S_n = \sum_{i=1}^n \xi_{q_1(i)} \cdots \xi_{q_\ell(i)}$ where now $p_n \equiv (\lambda/n)^{1/\ell}$ and $1 \leq q_1(i) < \cdots < q_\ell(i)$ are integer valued increasing functions. We obtain also Poissonian limit for numbers of arrivals to small sets of ℓ-tuples $X_{q_1(i)}, ..., X_{q_\ell(i)}$ for some Markov chains X_n and for numbers of arrivals of $T^{q_1(i)}x, ..., T^{q_\ell(i)}x$ to small cylinder sets for typical points x of a subshift of finite type T.

1. Introduction

The classical Poisson limit theorem taught in the first probability course says that if

$$\lim_{n\to\infty} np_n = \lambda > 0,\ p_n > 0 \tag{1.1}$$

then the binomial distribution with parameters (n, p_n) converges as $n \to \infty$ to the Poisson distribution with parameter λ. In other words, if $\xi_1^{(n)}, \xi_2^{(n)}, \xi_3^{(n)} ...,\ n = 1, 2, ...$ is an array of independent Bernoulli random variables satisfying

$$p_n = P\{\xi_i^{(n)} = 1\} = 1 - P\{\xi_i^{(n)} = 0\} \tag{1.2}$$

and (1.1) holds true then the sum

$$S_n = \sum_{l=1}^n \xi_l^{(n)} \tag{1.3}$$

converges in distribution to a Poisson random variable with the parameter λ.

It turns out that assuming

$$\lim_{n\to\infty} np_n^\ell = \lambda \tag{1.4}$$

the above result can be extended to "nonconventional" sums of the form

$$S_n = \sum_{l=1}^n \xi_{q_1(l)}^{(n)} \xi_{q_2(l)}^{(n)} \cdots \xi_{q_\ell(l)}^{(n)} \tag{1.5}$$

Date: October 11, 2011.
2000 Mathematics Subject Classification. Primary: 60F05 Secondary: 37D35, 60J05.
Key words and phrases. Poisson limit theorem, Markov chain, subshift of finite type.

where $l \leq q_1(l) < q_2(l) < ... < q_\ell(l)$ are increasing functions taking on integer values on integers. The name "nonconventional" comes from [10] where ergodic theorems for sums of the form (1.5) were studied. Recently, strong laws of large numbers and central limit theorems type results were obtained in [13] and [14] even for more general expressions.

We will consider also a nonconventional Poisson limit theorem for sums of the form

$$S_n = \sum_{l=1}^{n} \prod_{j=1}^{\ell} \mathbb{I}_{\Gamma_n}(X_{q_j(l)}) \tag{1.6}$$

(where $\mathbb{I}_\Gamma$ is the indicator of a set Γ), which counts the number of arrivals by a Markov chain $X_0, X_1, ...$ to small sets Γ_n at all times $q_j(l)$, $j = 1, ..., \ell$ when l runs from 1 to n. We suppose that the Markov chain has bounded transition densities which satisfy a Doeblin type condition. This ensures existence of a unique invariant (probability) measure μ. Assuming that

$$\lim_{n\to\infty} n(\mu(\Gamma_n))^\ell = \lambda > 0 \tag{1.7}$$

we show that S_n converges in distribution to a Poisson random variable provided $q_{i+1}(l) - q_i(l) \to \infty$, $l = 1, ..., \ell - 1$ as $l \to \infty$.

Another nonconventional Poisson limit theorem we deal with in this paper concerns multiple arrivals to shrinking cylinder sets by subshifts. Namely, we show that the expressions of the form

$$S_n = \sum_{l=1}^{\lambda/(P(B_n))^\ell} \prod_{j=1}^{\ell} \mathbb{I}_{B_n} \circ T^{q_j(l)} \tag{1.8}$$

have almost surely asymptotically in n the Poisson distribution with a parameter λ provided

$$\lim_{n\to\infty} n(P(B_n))^\ell = \lambda \tag{1.9}$$

where B_n's are certain cylinder sets of the subshift space Ω, P is a Gibbs measure, T is a left shift and $l \leq q_j(l)$, $j = 1, ..., \ell$ are increasing integer valued functions with sufficiently fast growth of differences $q_{j+1}(l) - q_j(l)$ as $l \to \infty$. This assertion generalizes in the nonconventional direction the results from [15] and [6]. We mention also a related result in the conventional setup about Poisson limits for numbers of arrivals to small shrinking sets by hyperbolic dynamical systems obtained in [7].

Observe that if

$$\tau_{B_n}(\omega) = \min\{l : T^{q_j(l)}\omega \in B_n, \forall j = 1, ..., \ell\}$$

then the above result yields that asymptotically the distribution of $(P(B_n))^\ell \tau_{B_n}$ is exponential since

$$P\{(P(B_n))^\ell \tau_{B_n} > \lambda\} = P\{S_n = 0\} \to e^{-\lambda} \text{ as } n \to \infty.$$

More advanced results about limiting exponential distributions of properly normalized first return times to small shrinking sets (see, for instance, [1], [2] and references there) and about Poissonian asymptotical behavior of distributions of numbers of arrivals to small shrinking sets (see, for instance, [11], [12] and references there) should be possible to derive in the nonconventional framework, as well.

2. Preliminaries and main results

We start with a probability space $(\Omega, \mathcal{F}, P)$ and an array of independent Bernoulli 0–1 random variables $\xi_1^{(n)}, \xi_2^{(n)}, ..., n = 1, 2, ...$ taking on value 1 with probability $p_n \in (0, 1)$. Our setup includes also increasing functions $l \le q_1(l) < q_2(l) < ... < q_\ell(l)$ taking on integer values on the integers. For any set $\Gamma \subset \mathbb{Z}_+$ of nonnegative integers put

$$P_\lambda(\Gamma) = \sum_{l \in \Gamma} e^{-\lambda} \frac{\lambda^l}{l!} \tag{2.1}$$

which is the probability assigned to Γ by the Poisson distribution with a parameter $\lambda > 0$.

2.1. **Theorem.** *Set $\lambda_n = np_n^\ell$. Then*

$$\sup_{\Gamma \subset \mathbb{Z}_+} |P\{S_n \in \Gamma\} - P_\lambda(\Gamma)| \le (2\ell^2 + 1)p_n + 2|\lambda - \lambda_n| e^{\max(\lambda, \lambda_n)}, \tag{2.2}$$

where S_n is defined by (1.5), and the right hand side of (2.2) tends to zero provided (1.4) holds true.

We will prove this assertion in Section 3 relying on Poisson approximation results for families of dissociated random variables from [4] and [5]. We will show also that the convergence of $P\{S_n \in \Gamma\}$ to $P_\lambda(\Gamma)$ can be derived relying on the result of [16], as well, though this does not give the speed of convergence as in (2.2).

As a natural intermediate step from the independent case to a stationary (dynamical systems) case we exhibit a nonconventional Poisson limit theorem for Markov chains. Namely, let $X_0, X_1, ...$ be a Markov chain on a measurable state space $(M, \mathcal{B})$ whose one-step and n-step transition probabilities $P(x, \cdot)$ and $P(n, x, \cdot)$ satisfy the conditions

$$P(x, \Gamma) \le Cm(\Gamma) \text{ and } P(n_0, x, \Gamma) \ge C^{-1} m(\Gamma) \;\; \forall \Gamma \subset M, \, \Gamma \in \mathcal{B} \tag{2.3}$$

for some probability measure m on $(M, \mathcal{B})$, an integer $n_0 \ge 1$ and a constant $C > 0$. Denote by P_x, $x \in M$ the probability on the path space provided $X_0 = x$ and let ν be an arbitrary initial distribution (a probability measure on M). The Markov chain $X_0, X_1, X_2, ...$ will be considered now with respect to the probability $P_\nu = \int_M P_x d\nu(x)$ on the path space. Assume also that $l \le q_1(l) < q_2(l) < ... < q_\ell(l)$ is a sequence of integer valued increasing functions such that

$$\lim_{l \to \infty} (q_{i+1}(l) - q_i(l)) = \infty \;\; \forall i = 1, 2, ..., \ell - 1. \tag{2.4}$$

Let μ be the unique invariant measure of the Markov chain above which exists in view of the (strong) Doeblin condition (2.3) (see, for instance, [8], §5, Ch.V).

2.2. **Theorem.** *Let $\Gamma_n \in \mathcal{B}$ be a sequence of measurable subsets of M such that (1.7) holds true and let P_ν be a probability on the path space Ω of the Markov chain corresponding to any initial distribution ν. Then S_n defined by (1.6) converges in distribution as $n \to \infty$ on the probability space (Ω, P) to a Poisson random variable with the parameter λ.*

Next, we consider another setup where Ω is a space of sequences determined by a 0–1 matrix $A = (\alpha_{ij}, \, 1 \le i, j \le \iota)$, namely

$$\Omega = \{\omega = (\omega_0, \omega_1, \omega_2, ...) : \, 1 \le \omega_i \le \iota \text{ and } \alpha_{\omega_i \omega_{i+1}} = 1 \; \forall i \ge 0\}. \tag{2.5}$$

The space Ω together with the left shift T acting by $(T\omega)_i = \omega_{i+1}$ is called a subshift of finite type ([3]). We assume that $A^{\wp}$ is a positive matrix for some $\wp > 0$ which makes T topologically mixing. Let ϕ be a Hölder continuous function on Ω with respect to the metric

$$d(\omega, \tilde{\omega}) = \exp(-\min\{i \geq 0 : \omega_i \neq \tilde{\omega}_i\}).$$

There exists a unique T-invariant Gibbs probability P corresponding to ϕ (see [3]) characterized by the property that for some Π (called the topological pressure), any cylinder set $[a_0, a_1, ..., a_{n-1}] = \{\omega \in \Omega : \omega_i = a_i, i = 0, 1, ..., n-1\}$ and each $\omega \in [a_0, a_1, ..., a_{n-1}]$,

$$C^{-1} \leq \frac{P([a_0, a_1, ..., a_{n-1}])}{\exp(-\Pi n + \sum_{i=0}^{n-1} \phi \circ T^i(\omega))} \leq C \tag{2.6}$$

where $C > 0$ depends only on ϕ. Denote by $\mathcal{F}_n$ the (finite) σ-algebra generated by all cylinder sets $[a_0, a_1, ..., a_{n-1}] \subset \Omega$ and for each $\omega^* = (\omega_0^*, \omega_1^*, ...) \in \Omega$ set $C_n(\omega^*) = [\omega_0^*, \omega_1^*, ..., \omega_{n-1}^*]$. Next, let $l \leq q_1(l) < q_2(l) < ... < q_\ell(l)$ be a sequence of increasing functions taking on integer values on integers and such that for some $c, \gamma > 0$ and all $l \geq 1$,

$$q_{i+1}(l) - q_i(l) \geq c(\ln l)^{1+\gamma},\ i = 1, 2, ..., \ell - 1. \tag{2.7}$$

2.3. **Theorem.** *For some $s \geq 0$ and each $\omega^* \in \Omega$ let $B_n(\omega^*) \subset C_n(\omega^*)$, $B_n(\omega^*) \in \mathcal{F}_{n+[s \ln n]}$, $n = 1, 2, ...$ be arbitrary sequences of sets. For each sequence $N_n(\omega^*)$, $n = 1, 2, ...$ satisfying*

$$\lim_{n\to\infty} N_n(\omega^*)\big(P(B_n(\omega^*))\big)^\ell = \lambda > 0 \tag{2.8}$$

set

$$S_{n,\omega^*}(\omega) = \sum_{l=0}^{N_n(\omega^*)} \prod_{j=1}^{\ell} \mathbb{I}_{B_n(\omega^*)} \circ T^{q_j(l)}(\omega).$$

Then for P-almost all $\omega^* \in \Omega$,

$$\lim_{n\to\infty} P\{\omega \in \Omega : S_{n,\omega^*}(\omega) = k\} = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{2.9}$$

We will prove this assertion modifying appropriately the technique from [15] and [6] while relying on the basic result from [16] which became a major tool for deriving Poissonian type limit theorems in dynamical systems.

2.4. **Remark.** In Theorems 2.2 and 2.3 we rely on [16] which does not provide speed of convergence to the Poisson distribution as in Theorem 2.1. Still, verifying conditions of [16] we obtain certain estimates for speed of convergence of relevant quantities there so relying on the quantitive version of [16] obtained in [11] we can obtain some estimates on speed of convergence in Theorems 2.2 and 2.3, as well.

2.5. **Remark.** It is well known that already in the conventional $\ell = 1$ setup the assertion of Theorem 2.3 holds true not for all ω^* but only for almost all ω^* and a corresponding example appears already in [15]. In fact, when ω^* is a periodic point then already in the conventional $\ell = 1$ setup the distribution of S_{n,ω^*} from Theorem 2.3 will converge to a compound Poisson distribution (see [12]) and not to a Poisson one. The nonconvergence to a Poisson distribution is usually easy to see in this case checking that second moments do not converge to second moments of a

Poisson distribution because of "short returns" which will be excluded in our situation by the assumption (5.4) from Section 5. For Harris recurrent Markov chains convergence in distribution to compound Poisson random variables of their number of arrivals to small sets were studied in [9] (see also references there). It seems that many of the results about convergence to compound Poisson distributions can be extended to the nonconventional setup in the spirit of the present paper. This can be done not relying on the basic result from [16] but by proving directly convergence of moments employing combinatorial arguments similar to ones in the proofs below.

3. A nonconventional Poisson theorem

For any ordered collection of ℓ indices $J=(j_1,...,j_\ell)\in\mathbb{Z}_+^\ell$, $j_1<...<j_\ell$ set

$$X_J=\begin{cases}\xi_{q_1(i)}\xi_{q_2(i)}\cdots\xi_{q_\ell(i)} \text{ if } j_l=q_l(i) \text{ for } i\le n \text{ and all } l=1,...,\ell\\ 0 \text{ if there is no } i\ge 1,\, i\le n \text{ such that } j_l=q_l(i)\ \forall l=1,...,\ell.\end{cases}\tag{3.1}$$

The collection $\{X_J\}$ is an example of a so called dissociated family of random variables which means that if $\{X_J\}_{J\in\mathcal{J}}$ and $\{X_K\}_{K\in\mathcal{K}}$ are two subfamilies and $(\cup_{J\in\mathcal{J}}J)\cap(\cup_{K\in\mathcal{K}}K)=\emptyset$ then these subfamilies are independent. Thus, we can apply Theorem 2 from [4] (see also Section 2.3 and 9.3 in [5]) which yields that

$$\sup_{\Gamma\subset\mathbb{Z}_+}|P\{S_n\in\Gamma\}-P_{\lambda_n}(\Gamma)|\le\ min(1,\lambda_n^{-1})\big(I_1(n)+I_2(n)+I_3(n)\big)\tag{3.2}$$

where S_n is defined by (1.5), $\lambda_n=np_n^\ell$,

$$I_1(n)=\sum_J p_J^2,\ I_2(n)=\sum_J\sum_{K\ne J,\,K\cap J\ne\emptyset}p_Jp_K,\ I_3(n)=\sum_J\sum_{K\ne J,\,K\cap J\ne\emptyset}EX_JX_K$$

and

$$p_{(j_1,...,j_\ell)}=\begin{cases}p_n^\ell \text{ if } j_l=q_l(i)\ \forall l=1,...,\ell\\ 0 \text{ if there is no } i\ge 1 \text{ such that } j_l=q_l(i)\ \forall l=1,...,\ell.\end{cases}\tag{3.3}$$

Clearly,

$$I_1(n)=np_n^{2\ell}=p_n^\ell\lambda_n.\tag{3.4}$$

Observe that if $J=(j_1,...,j_\ell)$, $p_J\ne 0$ is fixed and $K=(k_1,...,k_\ell)$, $p_K\ne 0$ satisfies $K\cap J\supset\{j_{l_1}\}=\{k_{l_2}\}$ then such K is uniquely determined by l_1 and l_2. Hence,

$$I_2(n)\le n\ell^2p_n^{2\ell}=\ell^2p_n^\ell\lambda_n.\tag{3.5}$$

Next, if $K\ne J$ then $EX_JX_K\le p_n^{\ell+1}$ and since by the above argument there are no more than $n\ell^2$ terms in the sum for $I_3(n)$ we obtain that

$$I_3(n)\le n\ell^2p_n^{\ell+1}=\ell^2p_n\lambda_n.\tag{3.6}$$

It follows from (3.2)–(3.6) that

$$|P\{S_n\in\Gamma\}-P_{\lambda_n}(\Gamma)|\le(2\ell^2+1)p_n$$

which yields (2.2) taking into account that

$$|P_\lambda(\Gamma)-P_{\lambda_n}(\Gamma)|\le 2|\lambda-\lambda_n|e^{\max(\lambda,\lambda_n)}.$$

□

Next, we formulate the main result from [16] which we will rely upon in Sections 4 and 5 but also, as a warm up, we will use it below in the simpler situation of this section in order to obtain an alternative proof of convergence of $P\{S_n\in A\}$

to $P_\lambda(A)$ as $n\to\infty$ though without an error estimate as in the right hand side of (2.2).

3.1. **Theorem.** *([16]) Let $\eta_1^{(n)},...,\eta_n^{(n)}$, $n=1,2,...$ be an array of 0–1 random variables, $J_r(n)$, $r\leq n$ be the family of all r-tuples $(i_1,i_2,...,i_r)$ of mutually distinct indices between 1 and n and for any $(i_1,...,i_r)\in J_r(n)$ set*

$$b^{(n)}_{i_1,...,i_r}=P\{\eta^{(n)}_{i_1}=...=\eta^{(n)}_{i_r}=1\}.$$

Assume that

$$\lim_{n\to\infty}\max_{1\leq i\leq n}b_i^{(n)}=0,\;\;\lim_{n\to\infty}\sum_{i=1}^n b_i^{(n)}=\lambda>0, \tag{3.7}$$

for $n=1,2,...$ there exist "rare" sets $I_r(n)\subset J_r(n)$ such that

$$\lim_{n\to\infty}\sum_{(i_1,...,i_r)\in I_r(n)}b^{(n)}_{i_1...i_r}=\lim_{n\to\infty}\sum_{(i_1,...,i_r)\in I_r(n)}b^{(n)}_{i_1}\cdots b_{i_r}=0 \tag{3.8}$$

and uniformly in $(i_1,...,i_r)\in J_r(n)\setminus I_r(n)$,

$$\lim_{n\to\infty}\frac{b^{(n)}_{i_1...i_r}}{b^{(n)}_{i_1}\cdots b^{(n)}_{i_r}}=1. \tag{3.9}$$

Then for $S_n=\sum_{i=1}^n\eta_i^{(n)}$,

$$\lim_{n\to\infty}P\{S_n=k\}=\frac{\lambda^ke^{-\lambda}}{k!},\;k=0,1,2,.... \tag{3.10}$$

Now we set

$$\eta_i^{(n)}=\xi^{(n)}_{q_1(i)}\xi^{(n)}_{q_2(i)}\cdots\xi^{(n)}_{q_1(i)}$$

and check the conditions of Theorem 3.1. For any two positive integers $l,\tilde l$ set

$$\rho(l,\tilde l)=\min_{1\leq i,j\leq \ell}|q_i(l)-q_j(\tilde l)|. \tag{3.11}$$

A sequence $J=\{j_1,j_2,...,j_l\}$ of distinct positive integers will be called a cluster here if for any $j,\tilde j\in J$ there exists a chain $j_{i_1}=j,j_{i_2},...,j_{i_{m-1}},j_{i_m}=\tilde j$ of integers from J such that $\rho(j_{i_k},j_{i_{k+1}})=0$ for all $k=1,2,...,m-1$. Suppose that J is a part of another finite sequence $\tilde J$ of distinct positive integers then we say that J is a maximal cluster in $\tilde J$ if $J\cup\{\tilde j\}$ is already not a cluster for any $\tilde j\in\tilde J$. In the notations of Theorem 3.1 we define now "rare" sets $I_r(n)$ by $I_r(n)=\cup_{1\leq k\leq r-1}I_r^{(k)}(n)$ where $I_r^{(k)}(n)$ is the collection of all r-tuples from $J_r(n)$ which contain exactly k maximal clusters . Hence, $J_r(n)\setminus I_r(n)$ consists of r-tuples whose all maxinal clusters are singeltons.

Clearly,

$$b_i^{(n)}=p_i^{(n)}\to 0\text{ and }\sum_{i=1}^n b_i^{(n)}=np_n^\ell=\lambda_n\to\lambda\text{ as }n\to\infty,$$

and so (3.7) holds true. If $(i_1,...,i_r)\in J_r(n)\setminus I_r(n)$ then $b_{i_1...i_r}=p_n^{\ell r}=b_{i_1}\cdots b_{i_r}$, and so (3.9) is satisfied, as well. Next, if l is fixed and we know that $q_i(l)=q_j(m)$ then m is uniquely determined by l,i and j. Hence, if l is fixed and $\rho(l,m)=0$ then there exist no more than ℓ^2 possibilities for m. It follows that there are no more

than $r!n\ell^{2r}$ possibilities for the choice of numbers in any cluster in each sequence from $J_r(n)$. Hence,

$$\#(I_r^{(k)}(n)) \leq (r!n\ell^{2r})^k,\ k=1,...,r-1 \tag{3.12}$$

where $\#\Gamma$ denotes the cardinality of a set Γ.

Now, observe that for any sequence $(i_1,...,i_l)$, $l\geq 2$ of distinct indices, in particular, for a cluster $b_{i_1...i_l}\leq p_n^{\ell+1}$, and so

$$b_{i_1...i_r}\leq p_n^{k\ell+1} \text{ for any } (i_1,...,i_r)\in I_r^{(k)}(n),\ k\leq r-1 \tag{3.13}$$

since each $I_r^{(k)}(n)$ with $k\leq r-1$ contains at least one cluster which is not a singelton. Hence, by (3.12) and (3.13),

$$\sum_{(i_1,...,i_r)\in I_r(n)} b_{i_1...i_r}\leq p_n\sum_{k=1}^{r-1}\lambda_n^k(r!\ell^{2r})^k\to 0 \text{ as } n\to\infty \tag{3.14}$$

while

$$\sum_{(i_1,...,i_r)\in I_r(n)} b_{i_1}\cdots b_{i_r}\leq \sum_{k=1}^{r-1}\lambda_n^k p_n^{(r-k)\ell}(r!\ell^{2r})^k\to 0 \text{ as } n\to\infty \tag{3.15}$$

implying (3.8) and completing the proof of (3.10). □

4. Poisson limits for arrivals to small sets: Markov chains

Next, we prove Theorem 2.2. Set

$$a(l)=\min(\ln l,\min_{1\leq i\leq \ell-1}(q_{i+1}(l)-q_i(l)).$$

Now a sequence $J=\{j_1,j_2,...,j_l\}$ of distinct positive integers will be called an (a,n)-cluster if for any $j,\tilde j\in J$ there exists a chain $j_{i_1}=j,j_{i_2},...,j_{i_{m-1}},j_{i_m}=\tilde j$ of integers from J such that

$$\rho(j_{i_k},j_{i_{k+1}})\leq a(n)\ \ \forall k=1,2,...,m-1 \tag{4.1}$$

with $\rho(\cdot,\cdot)$ defined by (3.11). The definition of maximal clusters remains as before while we define rare sets $I_r(n)$ as collections of r-tuples $J=(i_1,i_2,...,i_r)$ from $J_r(n)$ (which, recall, denotes the collection of all r-tuples of mutually distinct indices between 1 and n) which either contain a cluster containing more than one element or

$$i_{\min}(J)=\min_{1\leq l\leq r} i_l\leq a(n). \tag{4.2}$$

By (2.3) and the Radon–Nikodim theorem there exists a transition density $p(x,y)$ so that

$$P(x,\Gamma)=\int_\Gamma p(x,y)dm(y) \text{ and } C^{-1}\leq p(x,y)\leq C.$$

Let $p(n,x,y)$ denotes the n-step transition density so that

$$P(n,x,\Gamma)=P_x\{X_n\in\Gamma\}=P\{X_n\in\Gamma|X_0=x\}=\int_\Gamma p(n,x,y)dm(y).$$

We will rely on the well known fact (see, for instance, [8], §5 in Ch. V) that under (2.3) there exists a unique invariant measure μ (i.e. $\int_M d\mu(x)P(x,\Gamma)=\mu(\Gamma)$, $\forall\Gamma\in\mathcal{B}$) having a density $p(x)=\frac{d\mu(x)}{dm(x)}$ satisfying $C^{-1}\le p(x)\le C$ and

$$\sup_{x,y\in M}|p(n,x,y)-p(y)|\le C_1e^{-\beta n}\ \forall n\ge 1 \tag{4.3}$$

for some $C_1,\beta>0$ independent of n.

Set

$$\eta_l^{(n)}=\prod_{j=1}^{\ell}\mathbb{I}_{\Gamma_n}(X_{q_j(l)}),\ l=1,2,....$$

We start verifying conditions of Theorem 3.1 in our situation observing that by the Chapman–Kolmogorov formula

$$\begin{aligned}(4.4)\quad b_l^{(n)}=P_\nu\{\eta_l^{(n)}=1\}=\int_M d\nu(x)\int_{\Gamma_n}p(q_1(l),x,x_1)\int_{\Gamma_n}p(q_2(l)-q_1(l),x_1,x_2)\\ ...\int_{\Gamma_n}p(q_\ell(l)-q_{\ell-1}(l),x_{\ell-1},x_\ell)dm(x_1)...dm(x_\ell)\end{aligned}$$

where, recall, ν is the initial distribution of the Markov chain. Since $q_1(l)\ge l$ we obtain from (2.4), (4.3) and (4.4) that for any $l\ge 1$,

$$|b_l^{(n)}-(\mu(\Gamma_n))^\ell|\le C_2(\mu(\Gamma_n))^\ell\exp(-\beta a(l)) \tag{4.5}$$

for some $C_2>0$ independent of l. Hence, by (1.7) and (4.5),

$$\max_{1\le l\le n}b_l^{(n)}\le(C_2+1)(\mu(\Gamma_n))^\ell\to 0\text{ as }n\to\infty \tag{4.6}$$

and

$$|\sum_{l=1}^n b_l^{(n)}-n(\mu(\Gamma_n))^\ell|\le C_2((\mu(\Gamma_n))^\ell\sum_{l=1}^n e^{-\beta a(l)}\to 0\text{ as }n\to\infty, \tag{4.7}$$

and so (3.7) is satisfied.

Observe that since $p(x,y)\le C$ then $p(l,x,y)\le C$ for all l,x,y in view of the Chapman–Kolmogorov formula. Let $(i_1,...,i_l)$ be a sequence of distinct integers such that for some pairs $(m_1,i_{j_1}),...,(m_k,i_{j_k})$,

$$q_{m_1}(i_{j_1})<q_{m_2}(i_{j_2})<...<q_{m_k}(i_{j_k}) \tag{4.8}$$

where pairs are different but either i's or m's may repeat themselves. It follows by the Chapman–Kolmogorov formula that

$$\begin{aligned}(4.9)\quad b^{(n)}_{i_1...i_l}&=P_\nu\{\eta^{(n)}_{i_1}=1,...,\eta^{(n)}_{i_l}=1\}\\ &\le P_\nu\{X_{q_{m_1}}(i_{j_1})\in\Gamma_n,...,X_{q_{m_k}}(i_{j_k})\in\Gamma_n\}\\ &=\int_M d\nu(x)\int_{\Gamma_n}p(q_{m_1}(i_{j_1}),x,x_1)\int_{\Gamma_n}p(q_{m_2}(i_{j_2})-q_{m_1}(i_{j_1}),x_1,x_2)\\ &\quad...\int_{\Gamma_n}p(q_{m_k}(i_{j_k})-q_{m_{k-1}}(i_{j_{k-1}}),x_{k-1},x_k)dm(x_1)...dm(x_k)\\ &\le(Cm(\Gamma_n))^k\le(C^2\mu(\Gamma_n))^k.\end{aligned}$$

Next, we represent the rare sets $I_r(n)$ in the form

$$I_r(n)=\big(\cup_{1\le k\le r}I_r^{(k,1)}(n)\big)\cup\big(\cup_{k=1}^{r-1}I_r^{k,0}(n)\big) \tag{4.10}$$

where $I_r^{(k,1)}(n)$ and $I_r^{(k,1)}(n)$ are the sets of r-tuples from $J_r(n)$, $r\le n$ which contain exactly k maximal (a,n)-clusters and each $I_r^{(k,1)}(n)$ contains an (a,n)-cluster J with $i_{\min}(J)\le a(n)$ while $I_r^{(k,0)}(n)$ contains no such (a,n)-clusters. Clearly, no

r-tuple from $I_r(n)$ may contain more than one maximal (a,n)-cluster J satisfying $i_{\min}(J) \le a(n)$. Observe that $I_r^{(r,l)}(n)$, $l=0,1$ contains only singelton (a,n)-clusters, and so it is a subset of $I_r(n)$ only if $l=1$. In order to estimate the cardinality of rare sets observe that if l and m belong to a same (a,n)-cluster and l is fixed then there are no more than $2^r(a(n))^r\ell^{r+1}$ possibilities for m. It follows that there are no more than $n(a(n))^{r^2}2^{r^2}\ell^{r(r+1)}r!$ possibilities for the choice of numbers in any cluster in $J=(i_1,...,i_l)$ with $i_{\min}(J)>a(n)$ while if $i_{\min}(J)\le a(n)$ then there are no more than $(a(n))^{r^2+1}2^{r^2}\ell^{r(r+1)}r!$ such choices. Hence,

$$\#(I_r^{(k,l)}(n)) \le (2^{r^2}\ell^{r(r+1)}r!)^k(a(n))^{kr^2+l}n^{k-l},\ k=1,2,...,r;l=0,1. \tag{4.11}$$

where $\#\Gamma$ denotes cardinality of a set Γ.

Now observe that each cluster $J=(i_1,...,i_l)$ which is not a singleton yields at least $\ell+1$ pairs $(m_1,i_{j_1}),...,(m_{\ell+1},i_{j_{\ell+1}})$ satisfying (4.8). Hence, any $\tilde J = (\tilde i_1,...,\tilde i_r)\in I_r^{(k,l)}(n)$ with $k<r$ yields at least $k\ell+1$ pairs satisfying (4.8) while each r-tuple from $I_r^{(r,l)}(n)$ yields at least $r\ell$ such pairs. Since $a(n)\le \ln n$ then these arguments together with (1.7) and (4.9)–(4.11) yield that

$$\begin{aligned}\textstyle\sum_{(i_1,...,i_r)\in I_r(n)} b^{(n)}_{i_1...i_r} &\le \textstyle\sum_{k=1}^r \sum_{(i_1,...,i_r)\in I_r^{(k,1)}(n)} b^{(n)}_{i_1...i_r}\\ &+\textstyle\sum_{k=1}^{r-1}\sum_{(i_1,...,i_r)\in I_r^{(k,0)}(n)} b^{(n)}_{i_1...i_r} \le C_3\big(\sum_{k=1}^r((a(n))^{r^2k+1}(\mu(\Gamma_n))^\ell\lambda_n^{k-1}\\ &+\textstyle\sum_{k=1}^{r-1}(a(n))^{r^2k+1}\lambda_n^k\mu(\Gamma_n)\big)\to 0 \text{ as } n\to\infty\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}&\textstyle\sum_{(i_1,...,i_r)\in I_r(n)} b^{(n)}_{i_1}\cdots b^{(n)}_{i_r}\\ &\le \textstyle\sum_{k=1}^r\sum_{(i_1,...,i_r)\in I_r^{(k,1)}(n)} b^{(n)}_{i_1}\cdots b^{(n)}_{i_r}\\ &+\textstyle\sum_{k=1}^{r-1}\sum_{(i_1,...,i_r)\in I_r^{(k,0)}(n)} b^{(n)}_{i_1}\cdots b^{(n)}_{i_r}\\ &\le C_3\big(\textstyle\sum_{k=1}^r((a(n))^{r^2k+1}(\mu(\Gamma_n))^\ell\lambda_n^{k-1}\\ &+\textstyle\sum_{k=1}^{r-1}(a(n))^{r^2k+1}\lambda_n^k(\mu(\Gamma_n))^{(r-k)\ell}\to 0 \text{ as } n\to\infty\end{aligned} \tag{4.13}$$

where $\lambda_n=n(\mu(\Gamma_n))^\ell$ and $C_3>0$ does not depend on n, which gives (3.8).

Next, let $(i_1,...,i_r)\in J_r(n)\setminus I_r(n)$. Then there exist pairs $(m_1,i_{j_1}),(m_2,i_{j_2}),...,(m_{r\ell},i_{j_{r\ell}})$ such that

$$i_{j_1}\ge a(n) \text{ and } |q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})|\ge a(n) \text{ for } l=1,2,...,r\ell-1. \tag{4.14}$$

Employing again the Chapman–Kolmogorov formula together with (4.3) similarly to (4.5) we obtain that for such $(i_1,...,i_r)$,

$$|b^{(n)}_{i_1...i_r}-(\mu(\Gamma_n))^{r\ell}|\le C_4\exp(-\beta a(n))(\mu(\Gamma_n))^{r\ell} \tag{4.15}$$

for some $C_4>0$ independent of n. This together with (4.5) yields (3.9) and completes the proof of Theorem 2.2. $\square$

5. Nonconventional Poisson limits for subshifts

Before proving Theorem 2.3 itself we recall few basic facts about Gibbs measures for topologically mixing subshifts of finite type whose proofs can be found, for

instance, in Ch.1 of [3]. Namely, in addition to (2.6) we have that for P-almost all ω,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n}\phi\circ T^i(\omega)=\int_\Omega\phi dP=\Pi-h_P(T) \tag{5.1}$$

where $h_P(T)>0$ is the Kolmogorov–Sinai entropy of T with respect to P. It follows from (2.6) and (5.1) that for P-almost all $\omega=(\omega_0,\omega_1,...)$,

$$\lim_{n\to\infty}\frac{1}{n}\ln P([\omega_0,\omega_1...,\omega_{n-1}])=-h_P(T)<0. \tag{5.2}$$

Another important fact which we need is the exponentially fast ψ-mixing of such subshifts, namely, that there exist constants $\beta,C>0$ such that for any two cylinder sets $U=[a_0,a_1,...,a_l]\subset\Omega$ and $V=[b_0,b_1,...,b_m]\subset\Omega$,

$$|P(U\cap T^{-n}V)-P(U)P(V)|\le Ce^{-\beta(n-l)}P(U)P(V) \tag{5.3}$$

provided $n\ge l+1$.

Set $a(n)=[\ln^{1+\varepsilon}n]$, $\varepsilon\in(0,\gamma)$ where γ is the same as in (2.7). Denote by Ω^* the set of all $\omega^*\in\Omega$ such that

$$C_n(\omega^*)\cap T^{-i}C_n(\omega^*)=\emptyset\quad\text{for all}\quad i=1,2,...,a(n) \tag{5.4}$$

and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\phi\circ T^i(\omega_n)=\Pi-h_P(T)\quad\text{for any sequence}\quad\omega_n\in C_n(\omega^*). \tag{5.5}$$

It is clear that Ω^* is measurable and it was shown in [15] and [6] that $P(\Omega^*)=1$. For reader's convenience we recall the corresponding argument observing, first, that (5.5) follows from (5.1). As to (5.4) we note that $C_n\cap T^{-i}C_n\ne\emptyset$ for a cylinder set C_n if and only if it contains a periodic point of period i. The number of periodic points of period l grows with l exponentially, and so there are at most $g^{a(n)}$ periodic points of periods up to $a(n)$ for some $g>1$. Thus, there exist at most $g^{a(n)}$ cylinder sets C_n of length n such that $C_n\cap T^{-i}C_n\ne\emptyset$ for some $i\le a(n)$. On the other hand, relying on (5.2) we see that $P(C_n(\omega^*))$, and so also $g^{a(n)}P(C_n(\omega^*))$, decay exponentially fast in n for P-almost all ω^*. This together with the Borel–Cantelli lemma completes the argument.

Next, we return to our nonconventional setup. Put

$$\eta_l^{(n)}(\omega)=\prod_{j=1}^{\ell}\mathbb{I}_{B_n(\omega^*)}\circ T^{q_j(l)}(\omega),\ l=1,2,...,N_n(\omega^*)$$

where $\omega^*\in\Omega^*$, $N_n(\omega^*)$ was defined in Theorem 2.3, $B_n(\omega^*)\subset C_n(\omega^*)$ and $B_n(\omega^*)\in\mathcal{F}_{n+[s\ln n]}$, $n=1,2,\ldots$. Observe a slight change of notations here in comparison to Theorem 3.1 by writing $\eta_l^{(n)}$ in place of $\eta^{(N_n(\omega^*))}$ which would be unwieldy. We start verifying the conditions of Theorem 3.1 observing that

$$b_l^{(n)}=P\{\eta_l^{(n)}=1\}=P\big(\cap_{j=1}^{\ell}T^{(q_j(l)-q_1(l))}B_n(\omega^*)\big). \tag{5.6}$$

Taking into account (2.7) and applying (5.3) repeatedly we obtain that

$$\big|b_l^{(n)}-\big(P(B_n(\omega^*))\big)^\ell\big|\le D_1\exp(-\beta a(n))\big(P(B_n(\omega^*))\big)^\ell \tag{5.7}$$

for all $l \geq L(n)$ where

$$L(n) = \min\{k : c(\ln k)^{1+\gamma} > 2(n+a(n))\} = \big[\exp\big(\frac{2(n+a(n))}{c}\big)^{\frac{1}{1+\gamma}}\big] + 1 \tag{5.8}$$

grows in n subexponentially and $D_1 > 0$ does not depend on l, n and ω^*.

Since $P(B_n(\omega^*))$ decays in n exponentially fast we obtain from (2.8) and (5.6)–(5.8) that

$$\max_{1\leq l\leq N_n(\omega^*)} b_l^{(n)} \leq P(B_n(\omega^*)) \to 0 \text{ as } n \to \infty \tag{5.9}$$

and

$$\big|\sum_{l=1}^{N_n(\omega^*)} b_l^{(n)} - N_n(\omega^*)\big(P(B_n(\omega^*))\big)^{\ell}\big| \tag{5.10}$$
$$\leq 2L(N)P(B_n(\omega^*)) + D_1\big(P(B_n(\omega^*))\big)^{\ell} N_n(\omega^*)e^{-\beta a(n)} \to 0 \text{ as } n \to \infty$$

yielding (3.7).

We introduce again an (a,n)-cluster which is a sequence $J = \{j_1, j_2, ..., j_l\}$ of distinct positive integers such that for any $j, \tilde{j} \in J$ there exists a chain $j_{i_1} = j, j_{i_2}, ..., j_{i_{m-1}}, j_{i_m} = \tilde{j}$ of integers from J such that

$$\rho(j_{i_k}, j_{i_{k+1}}) \leq n + a(n) \quad \forall k = 1, 2, ..., m-1 \tag{5.11}$$

with $\rho(\cdot,\cdot)$ defined by (3.11). Denote by $J_r(n) = J_r(n,\omega^*)$ the set of all r-tuples $(i_1, ..., i_r)$ of distinct integers between 1 and $N_n(\omega^*)$. The definition of maximal clusters remains the same as before and we define rare sets $I_r(n) = I_r(n,\omega^*)$ as collections of r-tuples $J = (i_1, i_2, ..., i_r) \in J_r(n)$ which either contain not only singelton clusters or

$$i_{\min}(J) = \min_{1\leq j\leq r} i_j \leq L(n). \tag{5.12}$$

Next, we represent the rare sets $I_r(n)$ in the form

$$I_r(n) = \big(\cup_{1\leq l\leq k\leq r} I_r^{(k,l)}(n)\big) \cup \big(\cup_{k=1}^{r-1} I_r^{k,0}(n)\big) \tag{5.13}$$

where each $I_r^{(k,l)}(n) \subset J_r(n)$ contains exactly k maximal (a,n)-clusters while l of them are collections J satisfying $i_{\min}(J) \leq L(n)$. In order to estimate cardinality of $I_r^{(k,l)}(n)$ observe that if i_1 and i_2 belong to the same (a,n)-cluster and i_1 is fixed then there are no more than $2^r\ell^{r+1}(n+a(n))^r$ possibilities for i_2. If $i \leq L(n)$ then, of course, $L(n)$ bounds the number of choices for i. It follows that there are no more than $2^{r^2}N_n(\omega^*)\ell^{r(r+1)}r!(n+a(n))^{r^2}$ possibilities for the choice of integers in any cluster in $J = (i_1, ..., i_l)$ with $i_{\min}(J) > L(n)$ while if $i_{\min}(J) \leq L(n)$ then there are no more than $2^{r^2}L(n)\ell^{r(r+1)}r!(n+a(n))^{r^2}$ such choices. Thus, (4.11) takes here the form

$$\#(I_r^{(k,l)}(n)) \leq (2^{r^2}\ell^{r(r+1)}r!)^k(n+a(n))^{kr^2}(L(n))^l(N_n(\omega^*))^{k-l} \tag{5.14}$$

for $k = 1, 2, ..., r$ and $l = 0, 1, ..., k$.

Let $(i_1, ..., i_l)$ be a sequence of distinct integers such that for some pairs $(m_1, i_{j_1}), ..., (m_k, i_{j_k})$,

$$q_{m_1}(i_{j_1}) < q_{m_2}(i_{j_2}) - (n+a(n)) < q_{m_3}(i_{j_3}) - 2(n+a(n)) \tag{5.15}$$
$$< ... < q_{m_k}(i_{j_k}) - (k-1)(n+a(n))$$

where pairs are different but either i's or m's may repeat themselves. Then applying repeatedly (5.3) we obtain that for all n large enough (say, when $a(n) > s\ln n$),

$$b^{(n)}_{i_1\dots i_l} \le P\big(B_n(\omega^*)\cap T^{-(q_{m_2}(i_{j_2})-q_{m_1}(i_{j_1}))}B_n(\omega^*)\cap \dots\cap T^{-(q_{m_k}(i_{j_k})-q_{m_1}(i_{j_1}))}B_n(\omega^*)\big) \le D_2\big(P\big(B_n(\omega^*)\big)^k \tag{5.16}$$

for some $D_2>0$ independent of $n,\omega^*,i_1,...,i_l$ and $B_n(\omega^*)$.

By the definition each $J\in I_r^{(k,l)}(n)$ consists of k maximal (a,n)-clusters $J_1,...,J_l,...,J_k$ such that $i_{\min}(J_j)\le L(n)$ for $j=1,...,l$ and $i_{\min}(J_j)>L(n)$ for $j=l+1,...,k$. For each $j=1,...,k$ choose $\tilde i_j\in J_j$ arbitrarily and set $i_j=\tilde i_j$ for $j=1,...,l$ and $i_{l+(j-l)\ell+b}=\tilde i_j$ for $j=l,l+1,...,k-1$ and $b=1,...,\ell$. Set also $m_j=1$ for $j=1,...,l$ and $m_{l+(j-l)\ell+b}=b$ for $j=l,l+1,...,k-1$ and $b=1,...,\ell$. Next, we reorder the pairs (m_j,i_j) so that $q_{m_{j_b}}(i_{j_b})$ increases in b and it follows from the definition of $L(N)$ and of (a,n)-clusters that $l+(k-l)\ell$ pairs (m_{j_b},i_{j_b}), $b=1,...,l+(k-l)\ell$ will satisfy (5.15). This together with (5.16) yields that for any $(i_1,...,i_r)\in I_r^{(k,l)}(n)$ and n large enough,

$$b^{(n)}_{i_1\dots i_l}\le D_3\big(P(B_n(\omega^*))\big)^{(k-l)\ell+l} \tag{5.17}$$

for some $D_3>0$ independent on n,ω^* and $B_n(\omega^*)$.

The estimate (5.17) will suffice for our purposes when $l\ge 1$ but for $J=(\tilde i_1,...,\tilde i_r)\in I_r^{(k,0)}(n)$, $k<r$ a better estimate of $b^{(n)}_{\tilde i_1\dots\tilde i_l}$ will be needed. So let such J consists of k maximal (a,n)-clusters $J_1,...,J_k$ and since $k<r$ one of them must be not a singelton. Suppose, for instance, $i,\tilde i\in J_1$, $i\ne\tilde i$ and without loss of generality assume that

$$|q_j(i)-q_{\tilde j}(\tilde i)|\le n+a(n) \text{ for some } j,\tilde j=1,...,\ell.$$

Set $i_1=i$ and choose $i_b\in J_b, b=2,3,...,k$ arbitrarily. Now, order $k\ell+1$ pairs $(\tilde j,\tilde i)$ and $(l,i_b), l=1,...,\ell; b=1,...,k$ to obtain pairs $(m_1,i_{j_1}),(m_2,i_{j_2}),...,(m_{k\ell+1},i_{j_{k\ell+1}})$ so that $q_{m_b}(i_{j_b})$ is nondecreasing in b. Let $q_{m_l}(i_{j_l})=q_j(i)$ and assume without loss of generality that $q_{\tilde j}(\tilde i)\ge q_j(i)$. Then we must have $q_{\tilde j}(\tilde i)=q_{m_{l+1}}(i_{j_{l+1}})$ and

$$\begin{aligned}q_{m_1}(i_{j_1})&<q_{m_2}(i_{j_2})-(n+a(n))<...<q_{m_l}(i_{j_l})-(l-1)(n+a(n))\\&\le q_{m_{l+1}}(i_{j_{l+1}})-(l-1)(n+a(n))\\&<q_{m_{l+2}}(i_{j_{l+2}})-l(n+a(n))<...<q_{m_{k\ell+1}}(i_{j_{k\ell+1}})-(k\ell-1)(n+a(n)).\end{aligned}$$

Applying repeatedly (5.3) we obtain from here similarly to (5.16) that for all n large enough,

$$\begin{aligned}b_{\tilde i_1\dots\tilde i_r}&\le P\big(B_n(\omega^*)\cap T^{q_{m_2}(i_{j_2})-q_{m_1}(i_{j_1})}B_n(\omega^*)\cap\dots\\&\quad\dots\cap T^{q_{m_{k\ell}}(i_{j_{k\ell}})-q_{m_1}(i_{j_1})}B_n(\omega^*)\big)\\&\le D_4\big(P(B_n(\omega^*))\big)^{k\ell-1}P\big(B_n(\omega^*)\cap T^{q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})}B_n(\omega^*)\big)\end{aligned} \tag{5.18}$$

for some $D_4>0$ independent of n, ω^* and $B_n(\omega^*)$.

By (5.4) if

$$q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})\le a(n)$$

then

$$B_n(\omega^*)\cap T^{q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})}B_n(\omega^*)=\emptyset,$$

and so the right hand side of (5.18) is zero. If

$$n+[s\ln n]<q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})\le n+a(n)$$

then we still can use (5.3) to obtain that

$$P\big(B_n(\omega^*)\cap T^{q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})}B_n(\omega^*)\big)\le C\big(P(B_n(\omega^*))\big)^2. \tag{5.19}$$

Now let

$$a(n)<q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})\le n+[s\ln n]. \tag{5.20}$$

We can represent $B_n(\omega^*)$ as a disjoint union

$$B_n(\omega^*)=\cup_{b=1}^{m_n(\omega^*)}C^{(b)}_{n,s}(\omega^*)$$

where $C^{(b)}_{n,s}(\omega^*)$, $b=1,...,m_n(\omega^*)$ are cylinder sets of the length $n+[s\ln n]$ and $m_n(\omega^*)\le n^{\eta s}$ for some $\eta>0$. Employing (2.6) and (5.5) we obtain that under (5.20) for any $b,d=1,...,m_n(\omega^*)$,

$$P\big(C^{(b)}_{n,s}(\omega^*)\cap T^{q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})}C^{(d)}_{n,s}(\omega^*)\big)\le D_5P(C^{(b)}_{n,s}(\omega^*))\exp(-\delta a(n)) \tag{5.21}$$

for some $D_5,\delta>0$ where we can take $\delta=\frac12 h_P(T)$ provided n is large enough. Summing in b and d in (5.21) we obtain that

$$P\big(B_n(\omega^*)\cap T^{q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})}B_n(\omega^*)\big)\le D_5n^{\eta s}P(B_n(\omega^*))\exp(-\delta a(n)) \tag{5.22}$$

provided (5.20) holds true. In view of (2.6) and (5.2) the estimate (5.22) is weaker than (5.19) so we will use the former in both cases. Hence, from (5.18), (5.19) and (5.22) it follows that for any $(\tilde i_1,...,\tilde i_r)\in I_r^{(k,0)}(n)$ for all n large enough

$$b^{(n)}_{\tilde i_1,...,\tilde i_r}\le D_4D_5n^{\eta s}\big(P(B_n(\omega^*))\big)^{k\ell}\exp(-\delta a(n)). \tag{5.23}$$

Now (5.9), (5.14), (5.17), (5.23) together with the definition of $a(n)$ yield that

$$\begin{aligned}\textstyle\sum_{(i_1,...,i_r)\in I_r(n)}b^{(n)}_{i_1...i_r}&\le\textstyle\sum_{k=1}^r\sum_{l=1}^k\sum_{(i_1,...,i_r)\in I_r^{(k,l)}(n)}b^{(n)}_{i_1...i_r}\\&\quad+\textstyle\sum_{k=1}^{r-1}\sum_{(i_1,...,i_r)\in I_r^{(k,0)}(n)}b^{(n)}_{i_1...i_r}\\&\le D_6\big(\textstyle\sum_{k=1}^r(n+a(n))^{r^2k}(L(n))^l\big(P(B_n(\omega^*))\big)^l(\lambda_n(\omega^*))^{k-l}\\&\quad+\textstyle\sum_{k=1}^{r-1}(n+a(n))^{r^2k}(\lambda_n(\omega^*))^kn^{\eta s}\exp(-\delta a(n))\big)\to0\text{ as }n\to\infty\end{aligned} \tag{5.24}$$

and

$$\begin{aligned}&\textstyle\sum_{(i_1,...,i_r)\in I_r(n)}b^{(n)}_{i_1}\cdots b^{(n)}_{i_r}\\&\le\textstyle\sum_{k=1}^r\sum_{l=1}^k\sum_{(i_1,...,i_r)\in I_r^{(k,l)}(n)}b^{(n)}_{i_1}\cdots b^{(n)}_{i_r}\\&\quad+\textstyle\sum_{k=1}^{r-1}\sum_{(i_1,...,i_r)\in I_r^{(k,0)}(n)}b^{(n)}_{i_1}\cdots b^{(n)}_{i_r}\\&\le D_6\big(\textstyle\sum_{k=1}^r\sum_{l=1}^k((n+a(n))^{r^2k}(L(n))^\ell(\lambda_n(\omega^*))^{k-l}(P(B_n(\omega^*))^l\\&\quad+\textstyle\sum_{k=1}^{r-1}(n+a(n))^{r^2k}(\lambda_n(\omega^*))^k(P(B_n(\omega^*)))^{(r-k)\ell}\to0\text{ as }n\to\infty\end{aligned} \tag{5.25}$$

where $\lambda_n(\omega^*)=N_n(\omega^*)(P(B_n(\omega^*)))^{(r-k)\ell}$ and $D_6>0$ does not depend on n, which gives (3.8).

Next, let $(i_1,...,i_r)\in J_r(n)\setminus I_r(n)$. Then there exist pairs $(m_1,i_{j_1}),(m_2,i_{j_2}),...,(m_{r\ell},i_{j_{r\ell}})$ such that

$$i_{j_1}\ge L(n)\ \text{ and }\ q_{m_{l+1}}(i_{j_{l+1}})-q_{m_l}(i_{j_l})>n+a(n)\text{ for }l=1,2,...,r\ell-1. \tag{5.26}$$

Then applying repeatedly (5.3) we obtain similarly to (5.7) that for such $(i_1,...,i_r)$,

$$|b^{(n)}_{i_1...i_r} - \big(P(B_n(\omega^*))\big)^{r\ell}| \leq D_7 \exp(-\beta(a(n)-[s\ln n]))\big(P(B_n(\omega^*))\big)^{r\ell} \tag{5.27}$$

for some $D_7 > 0$ independent of n, ω^* and $B_n(\omega^*)$. This together with (5.7) yields (3.9) and completes the proof of Theorem 2.3. □

References

[1] M. Abadi, Sharp error terms and necessary conditions for exponential hitting times in mixing processes, Ann. Probab. 32, 243–264 (2004).

[2] M. Abadi and B. Saussol, Hitting and returning into rare events for all alpha-mixing processes, Stoch. Proc. Appl., to appear.

[3] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Math. 470, Springer–Verlag, Berlin, 1975.

[4] A.D. Barbour and G.K. Eagleson, *Poisson convergence for dissociated statistics*, J. R. Statist. Soc. B, 46, 397–402 (1984).

[5] A.D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Oxford Univ. Press, Oxford (1992).

[6] M. Denker, *Remarks on weak limit laws for fractal sets*, in: Fractal geometry and stochastics (C.Bandt, S.Graf& M.Zahle, eds.) Progress in Probab. 37, 167–178, Birkhäuser, Basel (1995).

[7] D. Dolgopyat, *Limit theorems for partially hyperbolic systems*, Trans. Amer. Math. Soc. 356, 1637–1689 (2004).

[8] J.L. Doob, *Stochastic Processes*, J.Wiley, New York, 1990.

[9] T. Erhardsson, *Compound Poisson approximation for Markov chains using Stein's method*, Ann. Probab. 27, 565–596 (1999).

[10] H. Furstenberg, *Nonconventional ergodic averages*, Proc. Symp. Pure Math. 50, 43–56 (1990).

[11] N. Haydn and S. Vaienti, *The limiting distribution and error terms for return times of dynamical systems*, Discr. Cont. Dyn. Sys. A, 10, 589–616 (2004).

[12] N. Haydn and S. Vaienti, *The compound Poisson distribution and return times in dynamical systems*, Probab. Th. Rel. Fields 144, 517–542 (2009).

[13] Yu. Kifer, *A nonconventional strong law of large numbers and fractal dimensions of some multiple recurrence sets*, Stoch. Dynam., to appear.

[14] Yu. Kifer and S.R.S. Varadhan, *Nonconventional limit theorems in discrete and continuous time via martingales*, Preprint, arXiv:1012.2223.

[15] B. Pitskel, *Poisson limit law for Markov chains*, Ergod. Th.& Dynam. Sys. 11, 501–513 (1991).

[16] B.A. Sevast'yanov, *Poisson limit law for a scheme of sums of dependent random variables*, Th. Probab. Appl. 17, 695–699 (1972).

Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel
E-mail address: kifer@math.huji.ac.il